

Validity & Reliability Statement

- For the standardized Spanish proficiency test, validity as well as reliability will be high since it is a well established test for the language.
- The pretest as well as post test interviews have been designed by the researchers of the study based on a few references so validity can be questionable. The reliability of these, is also not high since the sample size of the study is small and the authors cannot claim that these instruments will produce consistent and coherent results.
- The validity of the study cannot be established since it is based on reference from prior work but the methodology hasn't been thoroughly tested. The same can be said for the reliability. The authors attempted to mitigate effects or sources of bias.
- The validity of thematic analysis depends on several factors, including the research question, data collection and analysis methods, and the expertise of the researcher. The expertise of the researcher and personal bias could take away from the validity of this research.
- For the interviews validity would be high as the authors believe that they came up with interview questions relevant to the research questions.
- The inter-rater reliability would be low as due to the lack of time, as the other authors weren't able to cross-check the codebook rigorously
- The reliability of this research would be low owing to the small sample size and this being a constrained task of just learning a particular language

Possible sources of bias

- Individual Bias - Since the participants picked were from a very specific demographic i.e. young adults with a college education, it is very likely to have produced individual bias.
- Sequential Bias and Order Bias - Even though this was managed using a latin-square balance, there still exists some bias which cannot be mitigated.
- Learning Bias - There can be a learning bias causing the participants to perform better as the study went on since they would learn more and more and there can also be some overlap of information learned and later evaluated through the quiz.
- Technological Difficulty - The target demographic was fairly inexperienced with VR and learning through mobile applications is also not for everyone so this could have affected the motivation and overall learning experience of the participants.
- Motivation - This was a long study with a check in each week (more for the VR weeks) as well as a fair amount of time commitment and the participants were not compensated for their time either. This could have led to a loss of motivation which would cause participants to perform worse over time.

Discussion of results

- **Quantitative Analysis**

First the data was compiled into a single excel sheet extracted from the excel sheet where the data was recorded. We have one independent variable with two levels and a dependent variable. The independent variable here was medium and the dependent variable was the test scores. The first step was to check for normality. This was done through two methods. The first was plotting a histogram and a Q-Q plot. To further confirm the data was normal the Shapiro-Wilk normality test was performed. The p-value for the VR scores was 0.6644 and for the App scores was 0.1362. This evidently confirmed the data was normal. The data was then put through a two sample t-test since;

- Data was normally distributed
- Single independent variable with two levels
- Single dependent variable
- No significant outliers
- Variance for both levels of independent variable was the same

The test came out with a p-value of 0.3423 meaning the results were significant. The difference between the two means was 0.70834 (87.29167 for VR and 86.58333 for APP) showing there was no significant difference in the mean. The 95% confidence interval was between -0.7775497 and 2.1942163, which means the data could vary between that range with 95% confidence if this test was to be replicated. This also pointed towards there being no real difference between the two learning mediums. From these statistics we can conclude that there is no real difference when it comes to learning from different mediums with our configuration of the study. There is little evidence to support the alternate hypothesis 1.

However, when we see the difference between the final test scores and the initial test scores from the standardized Spanish proficiency test it can be seen that there is a 4-8% improvement for all participants. We can interpret that this platform was in fact helpful in learning Spanish to some degree which provides evidence supporting the second alternative hypothesis.

- **Qualitative Analysis Discussion**

All though our first hypothesis was rejected but one thing has been clear that over the course of 6 weeks participants showed an improvement in their Spanish proficiency. Be it proficiency in speaking or confidence in understanding and having day to day conversations.

The medium did not matter in terms of test scores or weekly quiz results but it did keep the participants engaged and on track. Most participants were excited about learning a new language with the use of virtual reality technology which contributed to their motivation and incentive of completing the study. The three themes which were identified in the qualitative analysis suggest that participants were more inclined to learn the language through virtual reality. Our second finding where we discuss the difficulties the participants had with the virtual reality hardware opens up further research into how those difficulties could be mitigated and how it could be made more accessible to people with visual impairments as small as having a myopic vision.

And lastly the authors show that the Learning platform which is Mondly app and VR helped the users keep track of their progress although there is scope for improvement in terms of navigation, field of view, orientation of the VR and graphics, it is a very good platform to learn a new language both on the mobile platform as well as the VR platform.

Contribution Statement











We divided the work in distinct parts and took ownership of each module.

Vamsi Munjuluri - Pre-test survey, post-test questionnaire (1), discussion, data collection (2 weeks)

Gautam Kushwah - Pre-test questionnaire, post-test questionnaire (4), qual analysis, discussion, data collection (2 weeks), study design(research questions, method, unit of analysis, concepts, instruments design)

Devansh Naimish Patel - post-test questionnaire (3), quant analysis, discussion, data collection (2 weeks), study design(hypothesis, method, validity and reliability statement, operationalization plan, instruments design)

External Links

1. [Working Document Link](#)
2. [R Code for Quant Analysis](#)
3. [Excel Sheet for collected data](#)
4.  Interview Script
5.  CodeBook CS594
6. Participants pre-test and post test interviews
 - 6.1.  P1
 - 6.2.  P2
 - 6.3.  P3
 - 6.4.  P4
 - 6.5.  P5
 - 6.6.  P6
 - 6.7.  P7
 - 6.8.  P8